ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202010.00073

The Impact of Misspecified Prior Information on Bayesian Estimation in Small-Sample Settings: A Hierarchical Model-Based Study

Authors: Shufang Zheng, Zhang Lijin, Pan Junhao, Pan Junhao

Date: 2020-10-27T00:00:00+00:00

Abstract

In research within the fields of psychology, education, and organizational behavior, researchers frequently encounter multilevel data with nested structures, such as subjects nested within communities, classes, clinics, etc. If the nested structure of the data is not taken into account, it may cause some statistical models to violate their independence assumptions, thereby producing substantial bias in the estimation of model parameters. Therefore, researchers often need to employ multilevel models to address issues arising from non-independent observations in multilevel data. However, due to practical constraints, multilevel data in actual research often exhibit small sample sizes at Level 1 or Level 2. The traditional frequentist Maximum Likelihood (ML) estimation method requires reliance on large samples and tends to encounter problems with parameter estimation and model convergence under small sample conditions. Bayesian estimation often offers greater advantages in small sample scenarios, but simultaneously, it is also more susceptible to the subjective specification of prior information. To investigate the potential negative effects of incorrect prior information in Bayesian estimation and compare them with traditional methods, this study, based on multilevel models and utilizing Monte Carlo simulation, examines the impact of prior information with varying information strength and degrees of bias on Bayesian estimation under different data types (dependent variables being continuous normally distributed data, continuous non-normal data, and dichotomous data), sample sizes, and intraclass correlation coefficients (ICC). Overall results demonstrate that prior distributions with means severely deviating from true values in Bayesian estimation can exert substantial negative impacts on parameter estimation, particularly when ICC is large, and when group sample sizes and prior distribution variance are small, and when the dependent variable is non-normally distributed or dichotomous, the negative effects of incorrect prior information become more pronounced. This study investigates the impact of incorrect prior information on Bayesian estimation in multilevel models and offers

recommendations, hoping to provide theoretical supplementation and empirical reference for the specification of prior distributions in Bayesian estimation.

Full Text

The Influence of Inaccurate Informative Priors on Bayesian Estimation in Small Samples: A Study Based on Multilevel Modeling

Zheng Shufang[1], Zhang Lijin[1], Pan Junhao[1] (Corresponding Author)

Abstract

In psychological, educational, and organizational behavior research, investigators frequently encounter multilevel data with nested structures (e.g., participants clustered within communities, classes, or clinics). Ignoring these hierarchical structures can violate the independence assumptions of statistical models, leading to substantially biased parameter estimates. Consequently, researchers often employ multilevel modeling to address problems arising from non-independent observations. However, due to objective constraints, real-world studies often feature small sample sizes at Level 1 and/or Level 2. Traditional frequentist maximum likelihood (ML) estimation, which relies on large-sample theory, frequently encounters issues with parameter estimation and model convergence in small-sample multilevel contexts. While Bayesian estimation offers advantages in small samples, it is simultaneously more vulnerable to subjectively specified prior information.

To investigate the potentially detrimental effects of inaccurate prior information on Bayesian approaches and compare their performance to traditional methods, we conducted Monte Carlo simulations within the multilevel modeling framework. We examined various conditions including different dependent variable types (continuous normal, continuous non-normal, and binary data), sample sizes, and intraclass correlation coefficients (ICCs). Overall, results revealed that prior distributions with means severely deviating from true values exert substantial negative impacts on Bayesian estimation, particularly when ICC is large, group sample sizes and prior variance are small, and when the dependent variable is non-normal or binary. This study investigates the influence of inaccurate priors on Bayesian estimation in multilevel models and offers recommendations, aiming to provide theoretical and empirical guidance for prior specification in Bayesian analysis.

Keywords: Multilevel Modeling; Small Samples; Bayesian Estimation; Inaccurate Prior Information

Introduction

In psychological, sociological, and organizational behavior research, studies frequently involve nested data structures, also known as multilevel data (Aryee et al., 2012; Holtmann et al., 2016; Hox et al., 2017). In such data, participants may come from different communities, classes, companies, or families. Because individuals within the same group tend to exhibit consistency, accounting for this internal coherence in modeling is essential (Sutton et al., 2013). When the nested structure of data is ignored, the dependency among individuals within groups can violate the independence assumptions of statistical models such as multiple linear regression (Cornfield, 1978; Rutterford et al., 2015; Van Breukelen & Candel, 2012), resulting in substantially biased parameter estimates (McNeish, 2016). To address statistical problems caused by non-independent observations within groups, researchers often need to employ multilevel modeling (Ryu, 2015).

The traditional estimation method for multilevel models is the frequentist approach (primarily maximum likelihood estimation, ML), which typically requires large sample sizes and often encounters convergence and parameter estimation problems when group sample sizes are small. However, collecting multilevel data often involves greater economic costs and higher difficulty, making it challenging to increase group sample sizes (Campbell & Walters, 2014). This problem is particularly acute in cluster randomized controlled trials (cRCTs), where small group sample sizes are common (McNeish, 2016). In psychology, researchers frequently design experimental studies to examine whether experimental manipulations or interventions produce significant effects by comparing experimental and control groups. Randomized controlled trials (RCTs) represent the gold standard for experimental research design (Campbell & Walters, 2014; Rutterford et al., 2015), yet many factors compel researchers to conduct interventions at the group rather than individual level. These factors may include geographical constraints, desire to reduce "treatment contamination" risk (Campbell, 2019), and aims to improve intervention efficiency, convenience, participant compliance, and reduce ethical concerns (Rutterford et al., 2015). Consequently, cRCT designs are common in psychological, sociological, and educational research, particularly in clinical psychology (Campbell, 2019; Ribeiro et al., 2018).

Given the limitations of traditional methods in small samples, researchers have proposed using Bayesian methods to address potential problems (Depaoli & van de Schoot, 2017; Kadane, 2015; McNeish, 2016; van de Schoot et al., 2014). Simulation studies have shown that in multilevel models, Bayesian methods require fewer groups than ML to obtain relatively accurate parameter estimates (Hox et al., 2012). Another advantage of Bayesian methods is their ability to incorporate prior information from previous research to improve estimation accuracy (Depaoli & Clifton, 2015; McNeish, 2016). However, due to lack of empirical research and theoretical guidance, prior distributions are difficult to specify perfectly, and inaccurate prior information may substantially impact parameter estimation.

Recent scholarship has begun examining how inaccurate priors affect specific models, including multitrait-multimethod (MTMM) models (Holtmann et al., 2016), latent growth models (Depaoli, 2014; Shi & Tong, 2017), and MIMIC models (Finch & Miller, 2019). Holtmann et al. (2016) compared different estimation methods in multilevel models under small-sample conditions within an MTMM framework. Their findings indicated that strongly inaccurate priors (i.e., small prior variance) could produce devastating results for parameter estimation, whereas weakly specified inaccurate priors (i.e., large prior variance) might outperform non-informative priors. However, Holtmann et al. (2016) only considered two-factor measurement models without covariates or regression paths, and assumed relatively large Level 2 sample sizes (50-200). In many multilevel studies (e.g., cRCTs), group (Level 2) sample sizes are typically smaller than 50, with within-group (Level 1) sample sizes generally not exceeding 100 (e.g., van der Putten et al., 2013; Ha et al., 2017; Newton et al., 2018; Shen et al., 2019).

Given the prevalence of cRCTs and current limitations in research on inaccurate priors, this study establishes a multilevel linear model in the cRCT context (McNeish, 2016) and employs Monte Carlo simulation. Because prior specification involves subjectivity, this simulation examines how inaccurate priors of varying informativeness and deviation from true values affect Bayesian estimation in multilevel linear models, investigating the robustness of Bayesian estimation to prior misspecification. Simultaneously, we compare Bayesian and traditional (ML) methods to help researchers select appropriate estimation methods in practice, hoping to provide guidance for method selection and prior specification in small-sample situations.

Multilevel Linear Model

Multilevel models analyze nested data structures to address statistical problems arising from non-independent observations within groups (Ryu, 2015). They decompose variance in outcome variables across different levels (e.g., individual, group, and higher levels), which can be explained by variables at each level (Heck & Thomas, 2015), thus flexibly accommodating interactions across levels.

For broad applicability, this study focuses on the most widely used multilevel linear model with random intercepts and slopes across groups. This model represents the standard cRCT framework (McNeish, 2016). The model equations and path diagram are as follows (McNeish, 2016):

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \beta_{3j}X_{3ij} + e_{ij} \quad (1)$$

$$\beta_{0j} = b_{00} + b_{01}W_{1j} + u_{0j} \quad (2)$$

$$\beta_{1j} = b_{10} + b_{11}W_{1j} + u_{1j} \quad (3)$$

$$\beta_{2j} = b_{20} + b_{21}W_{1j} + u_{2j} \quad (4)$$

$$\beta_{3j} = b_{30} + b_{31}W_{1j} + u_{3j} \quad (5)$$

[Figure 1: see original paper]

Where W_{1j} is a Level 2 predictor, typically a binary variable indicating experimental or control group in cRCTs; X_{1ij}-X_{3ij} are Level 1 predictors representing participant gender, age, pre-test scores, etc. β_{0j} is the random intercept, and β_{1j}-β_{3j} are random slopes, with subscript j indicating these intercepts and slopes may vary across groups. u_{0j}-u_{3j} are residual terms for β_{0j}-β_{3j}, respectively, following a multivariate normal distribution with mean 0 and variance-covariance matrix .

Traditionally, multilevel models employ frequentist methods for parameter estimation (Depaoli & Clifton, 2015). However, under ML theory, unbiased parameter estimation requires the assumption of asymptotic normality, demanding large sample sizes (McNeish, 2016). In multilevel analysis, Level 2 parameters are estimated by treating groups as units, so ML methods also require sufficient numbers of groups to satisfy asymptotic theory assumptions (Asparouhov & Muthén, 2010). Generally, as group sample size increases, the overall data become more normally distributed (Rutterford et al., 2015); thus, increasing the number of groups is more effective for increasing statistical power than increasing within-group sample size (Ribeiro et al., 2018). Numerous studies have noted that ML performs poorly in multilevel models with small total sample sizes (McNeish & Stapleton, 2014), particularly when both Level 1 and Level 2 sample sizes are small. Traditional ML methods often encounter convergence problems (Depaoli & Clifton, 2015; Hox & Maas, 2001; Hox et al., 2010; McNeish, 2016; Schoeneberger, 2015) and may produce unreliable, unstable parameter estimates (Asparouhov & Muthén, 2010; Hox & Maas, 2001; McNeish, 2016).

Bayesian Estimation

To address convergence and parameter estimation issues with traditional frequentist methods in small-sample multilevel models, researchers can employ Bayesian estimation (Depaoli & van de Schoot, 2017; Kadane, 2015; McNeish, 2016; van de Schoot et al., 2014). Generally, Bayesian estimation outperforms traditional methods in model convergence and parameter estimation under small-sample conditions (Lee & Song, 2004; Muthén & Asparouhov, 2012), especially when group sample sizes are small (Asparouhov & Muthén, 2010; Hox et al., 2012; Muthén & Asparouhov, 2012). Bayesian methods treat unknown

parameters as random variables, using prior information and likelihood functions to obtain posterior distributions that reflect the probability of parameters taking different values (Lynch, 2007). Unlike traditional frequentist approaches, Bayesian methods do not rely on large-sample asymptotic theory (McNeish, 2016). In Bayesian estimation, the posterior distribution of parameter values can be expressed as (Lynch, 2007):

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

Where y represents the vector of observed variables and θ represents the vector of unknown parameters. $p(\theta|y)$ denotes the posterior distribution, which depends on both prior information $p(\theta)$ and the data likelihood function $p(y|\theta)$ (Lynch, 2007). Generally, with large sample sizes, the posterior distribution is dominated by the likelihood function, while as sample size decreases, the influence of prior information on posterior estimation gradually increases. Therefore, prior specification requires greater caution in small-sample situations (Holtmann et al., 2016; McNeish, 2016).

Bayesian methods do not require assumptions about parameter sampling distributions or complex calculations to obtain standard errors, thus effectively overcoming convergence problems in small samples (Depaoli & Clifton, 2015; Levy & Choi, 2013). Another advantage is the ability to incorporate prior information from previous research to improve estimation accuracy (Depaoli & Clifton, 2015; McNeish, 2016). However, prior specification involves subjectivity, and due to insufficient empirical research and theoretical guidance, prior distributions may be misspecified, leading to inaccurate prior information.

To examine how prior informativeness (i.e., prior variance magnitude) affects Bayesian estimation, researchers have conducted simulation studies across various contexts (Depaoli & Clifton, 2015). For example, McNeish (2016) used cRCTs as a research context to build multilevel models, comparing different estimation methods including traditional frequentist ML, restricted maximum likelihood (REML), and Bayesian estimation, while examining how different levels of prior informativeness affected parameter estimation. Simulation results showed that prior informativeness substantially impacted parameter estimation, particularly with small total sample sizes and categorical dependent variables (Lynch, 2007; McNeish, 2016). However, like many simulation studies, McNeish (2016) investigated prior effects under the assumption of correctly specified priors, without considering the impact of inaccurate prior information. Generally, strongly inaccurate priors produce devastating results for parameter estimation, especially with small sample sizes. Yet some research suggests that weakly specified inaccurate priors (i.e., large prior variance) may outperform non-informative priors (Holtmann et al., 2016). Currently, many aspects of inaccurate prior effects in Bayesian estimation remain unclear and require further investigation (Finch & Miller, 2019; Holtmann et al., 2016). Therefore, this study conducts a series of simulations based on widely used multilevel linear models to explore

how inaccurate priors affect Bayesian estimation under different data conditions.

Study Design

Based on the multilevel linear model in Figure 1, this study examines the effects of different sample sizes (Level 1 and Level 2), intraclass correlation coefficients (ICC), and estimation methods. We generated simulated data using R and analyzed models with Mplus 8.0 (Muthén & Muthén, 1998-2017). For model simplification, the dependent variable follows a continuous normal distribution.

Using "cluster Randomised Controlled Trials" and "multilevel" as keywords, we searched Web of Science Core Collection for articles published in the past five years (2015-2019), retrieving and initially screening 95 relevant papers. Through abstract and full-text review, we coded the number of groups and average within-group sample sizes for studies reporting these details. As shown in [Figure 2: see original paper], the number of groups in recent cRCT research primarily ranges from 10-40, with most within-group sample sizes below 100. Therefore, based on literature review and coding results, along with recommendations from Depaoli and Clifton (2015), this study sets Level 2 sample sizes at 20, 30, 40, and 50, and Level 1 sample sizes at 30, 60, and 150.

In multilevel model research, investigators often examine how ICC at different levels may affect parameter estimation. ICC represents the proportion of total variance in a variable that can be explained by Level 2 variance, calculated as ICC = σ^2_B / (σ^2_B + σ^2_W), where σ^2_W and σ^2_B represent Level 1 and Level 2 variances, respectively. Larger ICC indicates greater between-group relative to within-group differences. Previous research shows that traditional frequentist methods may encounter parameter estimation and convergence issues when ICC is small (Hox & Maas, 2001; Preacher et al., 2011; Koch et al., 2015), while Bayesian methods may still produce accurate estimates (Hox et al., 2012; Muthén & Asparouhov, 2012). Simulation studies indicate that with non-informative priors, larger ICC may increase estimation bias for Level 2 parameters (Depaoli & Clifton, 2015). Additionally, in multilevel MIMIC models, Level 2 parameter estimation bias increases and statistical power decreases as ICC increases (Cao et al., 2019; Finch & French, 2011). Fang et al. (2019) suggested that future research should further examine how ICC affects Bayesian and other estimation methods across more multilevel models.

Generally, multilevel modeling is recommended when ICC $\geq$ 0.059 (Cohen, 1988). Previous multilevel simulation studies have primarily focused on ICC values between 0.05 and 0.2 (e.g., Depaoli & Clifton, 2015; Hox et al., 2010; Lüdtke et al., 2011; Preacher et al., 2011). Therefore, this study includes three ICC levels: 0.05, 0.1, and 0.2. To achieve these ICC levels, variances and residual variances for variables at each level were adjusted accordingly (Depaoli & Clifton, 2015).

The estimation methods include traditional frequentist ML^2 and Bayesian methods. Bayesian methods are divided into non-informative and informative priors.

For informative Bayesian estimation, prior distributions N(, σ^2) are specified for regression coefficients. The mean includes five levels: unbiased; offset left/right by 1 standard deviation (SD) from the true value; and offset left/right by 3 SDs. The variance σ^2 includes three levels: 10%, 20%, and 50% of the true regression coefficient value (Depaoli, 2014). Using b_{01} (true value = 1.5) as an example, [Figure 3: see original paper] illustrates the 15 different prior specifications varying in deviation and informativeness. Including the non-informative prior case (Mplus default: prior mean = 0, variance = ∞), Bayesian methods have $1 + 5 \times 3 = 16$ prior specifications. Combined with traditional ML estimation, 17 estimation methods are compared.

[Figure 3: see original paper]

In total, this study includes $3 \times 4 \times 3 \times 17 = 612$ simulation condition combinations. To reduce bias from random factors in model estimation, we generated 100 replications for each condition.

Results Evaluation Metrics

To evaluate parameter estimation performance across methods, we first examined convergence rates for each condition, then analyzed five metrics among converged results: relative bias of coefficient estimates, standard error ratio, mean square error (MSE), 95% confidence/credible interval coverage (95% CI coverage), and statistical power.

[2] Mplus uses robust maximum likelihood estimation (MLR) by default for multilevel models.

Given the simplicity of the multilevel linear model, all conditions achieved convergence. Since we only specified prior distributions for regression coefficients, evaluation focused exclusively on these parameters. To examine how different factors and their interactions affected each metric, we conducted ANOVA on evaluation metrics for each replication (note: for relative bias, absolute values were used in ANOVA to avoid cancellation of positive and negative deviations). As shown in , all evaluation metrics were affected to varying degrees by ICC, Level 1 and Level 2 sample sizes, prior variance magnitude, and mean deviation from true values, with potential interactions among factors.

In cRCTs, researchers typically focus on the effect of Level 2 predictor W_{1j} on the outcome, b_{01}. Therefore, to present results clearly, all subsequent analyses focus on b_{01}, examining its parameter estimation bias, standard error ratio, and other metrics.

Relative Bias

Relative bias measures the deviation of the mean parameter estimate from the true value across converged models, calculated as Relative Bias = ($\hat{\theta}$ -)/ , where

$\hat{\theta}$ is the average estimate and is the true parameter value. Relative bias within $±$10% is considered acceptable and unbiased (Flora & Curran, 2004).

Overall, relative bias patterns were consistent across parameters. Using b_{01} as an example ([Figure 4: see original paper]), ML and non-informative prior Bayesian estimation performed similarly, showing minimal influence from ICC levels. For informative prior Bayesian methods, parameter estimation bias was jointly affected by Level 2 sample size, prior variance, and mean deviation from true values, while Level 1 sample size showed minimal impact. Generally, as Level 2 sample size and prior variance increased, the effect of prior mean deviation on relative bias decreased. For the experimental effect (b_{01}) of primary interest in cRCTs, the negative impact of inaccurate priors on relative bias diminished as ICC decreased, though this ICC trend was not evident for other regression coefficients.

Standard Error Ratio

The standard error ratio is the ratio of the average standard error estimate to the empirical standard deviation across converged models. Accurate standard error estimation yields ratios approaching 1, with acceptable ranges typically between 0.9 and 1.1 (Cham et al., 2012).

Results for b_{01} standard error ratios are shown in [Figure 5: see original paper]. Informative prior Bayesian estimation generally overestimated standard error ratios beyond acceptable limits. As prior variance increased (weakening prior informativeness), standard error ratios decreased toward 1. Larger Level 2 sample sizes also substantially reduced standard error ratios in informative prior Bayesian estimation. When prior means deviated substantially from true values, standard error ratios were higher than with correctly specified priors under equivalent conditions. The negative impact of inaccurate priors on standard error ratios diminished as ICC decreased.

Mean Square Error

Mean square error (MSE) is the average squared error of parameter estimates across converged models, calculated as MSE = $\Sigma(\hat{\theta}_i$ - _i)^2 / number of converged models, where $\hat{\theta}_i$ is the estimate from the ith replication and _i is the true value. Smaller MSE indicates more accurate estimation.

As shown in [Figure 6: see original paper], with small samples, ML and non-informative prior Bayesian estimation tended to produce larger MSE. Bayesian estimation with correct priors performed optimally in terms of MSE, while inaccurate priors substantially increased MSE. As ICC decreased and Level 1/Level 2 sample sizes increased, the impact of prior mean deviation on MSE diminished. Level 1 sample size had weaker effects on MSE than Level 2 sample size. As Level 2 sample size increased, the amplifying effect of larger ICC on inaccurate priors' negative impact decreased. Additionally, larger prior variance weakened

the negative impact of inaccurate priors on MSE.

95% CI Coverage

The 95% CI coverage rate is the probability that the true parameter value falls within the 95% confidence or credible interval across converged replications. Frequentist confidence intervals rely on asymptotic theory, whereas Bayesian credible intervals are derived directly from posterior distribution percentiles without distributional assumptions. Accurate interval estimation should yield coverage near 0.95, with values above 0.9 generally considered acceptable (Collins et al., 2001).

As shown in [Figure 7: see original paper], 95% CI coverage generally increased with larger Level 2 sample sizes, while Level 1 sample size showed minimal effect. For informative prior Bayesian estimation, greater deviation of prior means from true values produced lower coverage rates. When prior mean deviation did not exceed 1 SD, 95% CI coverage remained acceptable (>90%). However, substantial deviation (3 SDs) substantially impacted coverage. The negative impact of inaccurate priors on 95% CI coverage decreased as ICC decreased and Level 2 sample size increased, though this ICC trend was not consistent for other regression coefficients.

Statistical Power

Statistical power is the probability of correctly rejecting a false null hypothesis—the proportion of converged models where the 95% confidence or credible interval for a non-zero parameter excludes zero. Power above 0.8 is generally acceptable (Cohen, 1988; Muthén & Muthén, 2002).

As shown in [Figure 8: see original paper], with small samples and large ICC, ML and non-informative prior Bayesian estimation showed low power. Bayesian estimation with correct priors maintained acceptable power (>0.8) even with small samples. Inaccurate priors that underestimated parameter effects reduced power, while those that overestimated effects showed higher power but potentially increased Type I error rates. The impact of inaccurate priors on power decreased as Level 2 sample size increased and ICC decreased.

Overall, Bayesian estimation with correct informative priors performed optimally across all metrics, particularly for parameter estimation bias. Non-informative prior Bayesian and ML methods showed similar patterns, performing poorly on MSE and power, especially with small samples and large ICC. When prior means deviated from true values, larger ICC and smaller Level 2 sample sizes exacerbated negative impacts. Stronger prior informativity also amplified the effects of inaccurate priors.

Discussion

In many psychological experiments (e.g., cRCTs), researchers must often build multilevel models due to nested data structures. These models frequently face insufficient sample sizes, creating substantial obstacles for data analysis. Bayesian estimation can incorporate prior information to improve parameter precision, but subjective prior specification may lead to inaccurate priors. Based on multilevel models, this study investigated how inaccurate priors affect Bayesian estimation across different sample sizes and ICC values, providing Mplus code and recommendations for prior specification. Inaccurate priors produced substantial bias when ICC was large and Level 1/Level 2 sample sizes were small. Stronger prior informativity amplified these negative effects. Therefore, researchers must be more vigilant about inaccurate priors and more cautious in prior specification when sample sizes are small and ICC is large.

Simulation results showed that larger Level 1 and Level 2 sample sizes reduced risks from inaccurate priors, consistent with Holtmann et al. (2016). Bayesian methods treat unknown parameters as random variables, using prior information and likelihood functions to obtain posterior distributions. With small samples, priors substantially influence posteriors, but this influence diminishes as sample size increases, with posterior distributions approximating the likelihood function (Lynch, 2007). Consequently, prior misspecification has more severe consequences with small samples, requiring greater caution.

For the primary effect of interest in cRCTs (b_{01}), inaccurate priors had stronger negative impacts when ICC was larger. This aligns with previous findings that in multilevel models, Level 2 effect estimation bias increases and power decreases as ICC increases (Cao et al., 2019; Finch & French, 2011). However, for other regression coefficients, ICC may have inconsistent effects on bias and power, consistent with Cao et al. (2019), who found that ICC did not significantly affect estimation or power for pure within-level effects or cross-level interactions. Future research should further examine how effects at different levels perform across ICC values and how inaccurate priors affect their estimation.

Results across different prior specifications showed that larger prior variance weakened the impact of inaccurate information. Even when priors deviated somewhat from true values, negative effects were well-controlled and diminished further with increased sample size. However, with small prior variance, even large sample sizes could not reduce the impact of inaccurate priors. Therefore, specifying priors with relatively large variance is more conservative.

Standard error ratio findings were consistent with previous research. As prior variance decreased, standard error ratios increased, indicating that informative Bayesian methods' posterior standard deviations tended to overestimate empirical standard deviations. This may occur because parameter estimation standard deviations decrease with prior variance, inflating the ratio (Holtmann et al., 2016). Particularly with inaccurate priors, Bayesian estimation tends to overestimate standard errors, so researchers should rely on credible intervals for

significance testing.

Limitations and Future Directions

As simulation studies cannot exhaust all possible conditions, this research focused on the most common sample size ranges in cRCTs. However, other sample size combinations may occur in practice. Moreover, this study only considered equal within-group sample sizes, whereas real-world studies often have unbalanced designs. Future research should examine whether unbalanced sample sizes exacerbate negative effects of inaccurate priors. Additionally, this study used large spans for Level 1 sample sizes; future studies could employ more fine-grained sample size specifications.

To simplify simulation conditions, we assumed normally distributed dependent variables. However, this assumption is frequently violated in empirical research. For example, variables may exhibit skewness due to researcher selection biases or ceiling/floor effects (Fleishman, 1978). Additionally, cRCTs may involve categorical outcome variables, particularly binary variables (Legare et al., 2015). Future research should explore how inaccurate priors affect estimation with different dependent variable types.

Furthermore, while this study focused on the intervention effect (b_{01}) of primary interest in cRCTs, ICC may have inconsistent effects on other regression coefficients. Future research should more deeply investigate how inaccurate priors affect parameters representing different level effects in multilevel models.

Conclusions

This study examined how priors of varying informativeness and deviation affect Bayesian estimation in multilevel linear models through Monte Carlo simulations. Discussing impacts of inaccurate priors on parameter estimation under different sample sizes, ICC values, and data types in cRCT contexts, we aim to provide theoretical and empirical guidance for prior specification. Our conclusions are:

1. In small-sample multilevel models, ML and non-informative prior Bayesian estimation may produce biased between-group parameter estimates. Therefore, with small samples, we recommend specifying informative priors for Bayesian methods whenever possible.

2. Severely inaccurate priors can have devastating effects on Bayesian parameter estimation, posing risks for empirical research. When prior means deviate from true values by no more than 1 SD, Bayesian estimation performance is generally acceptable.

3. When Level 2 sample sizes and prior variance are small, inaccurate priors substantially bias parameter estimation. For the intervention effect (b_{01})

of primary interest in cRCTs, larger ICC also amplifies negative effects of inaccurate priors. Therefore, greater caution is needed in these situations.

4. When prior information is insufficient and risks of inaccurate priors are high, researchers can mitigate negative effects by maximizing group sample sizes and increasing prior variance.

References

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis of latent variable models using Mplus (Technical report, Version 4). Retrieved from http://www.statmodel.com/download/BayesAdvantages18.pdf

Aryee, S., Walumbwa, F. O., Seidu, E. Y. M., & Otaye, L. E. (2012). Impact of high-performance work systems on individual- and branch- level performance: Test of a multilevel model of intermediate linkages. *Journal of Applied Psychology, 97*, 287–300.

Campbell, M. J. (2019). Cluster randomised trials. *Medical Journal of Australia, 210*(4).

Campbell, M.J., & Walters, S.J. (2014). *How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research*. Wiley.

Cao, C., Kim, E. S., Chen, Y.-H., Ferron, J., & Stark, S. (2019). Exploring the Test of Covariate Moderation Effects in Multilevel MIMIC Models. *Educational and Psychological Measurement, 79*(3), 512-544.

Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with nonnormal observed data: a comparison of four approaches. *Multivariate Behavioral Research, 47*(6), 840-876.

Cohen, John. (1988). *Statistical Power Analysis of the Behavioral Sciences* (2nd Editor). Hillside, N. J.: Eribaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330-351.

Cornfield, J. (1978). Symposium on chd prevention trials: design issues in testing life style intervention: randomization by group: a formal analysis. *American Journal of Epidemiology, 108*(2).

Depaoli, S. (2014). The Impact of Inaccurate "Informative" Priors for Growth Parameters in Bayesian Growth Mixture Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(2), 239-252.

Depaoli, S., & Clifton, J. P. (2015). A Bayesian Approach to Multilevel Structural Equation Modeling With Continuous and Dichotomous Outcomes. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(3), 327-351.

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods, 22*(2).

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in crosssectional multilevel models: A new look at an old issue. *Psychological Methods, 12*, 121–138.

Fang, J., Wen, Z. L., & Hau, K. T. (2019). Mediation Effects In 2-1-1 Multilevel Model: Evaluation Of Alternative Estimation Methods. *Structural Equation Modeling-a Multidisciplinary Journal, 26*(4), 591-606.

Finch, W. H., & French, B. F. (2011). Estimation of MIMIC model parameters with multilevel data. *Structural Equation Modeling, 18*, 229–

Finch, W. H., & Miller, J. E. (2019). The Use of Incorrect Informative Priors in the Estimation of MIMIC Model Parameters with Small Sample Sizes. *Structural Equation Modeling-a Multidisciplinary Journal, 26*(4), 497-508.

Fleishman, A. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*(4), 521-532.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods, 9*(4), 466.

Ha, A. S., Lonsdale, C., Ng, J. Y. Y., & Lubans, D. R. (2017). A school-based rope skipping program for adolescents: Results of a randomized trial. *Preventive Medicine, 101*, 188-194.

Heck, R. H. & Thomas, S. L. (2015). *An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches Using Mplus*, Routledge.

Holtmann, J., Koch, T., Lochner, K., & Eid, M. (2016). A Comparison of ML, WLSMV, and Bayesian Methods for Multilevel Structural Equation Models in Small Samples: A Simulation Study. *Multivariate Behavioral Research, 51*(5), 661-680.

Hox, J. J., & Maas, C. J. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(2), 157–174.

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica, 64*, 157–170.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Hox, J. J., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods, 6*(2), 87-93.

Kadane, J. B. (2015). Bayesian methods for prevention research. *Prevention Science, 16*, 1017-1025.

Koch, T., Schultze, M., Burrus, J., Roberts, R. D., & Eid, M. (2015). A multilevel CFA-MTMM model for nested structurally different methods. *Journal of Educational and Behavioral Statistics, 40*(5), 477-510.

Lee, S.-Y., & Song, X.-Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research, 39*(4), 653-686.

Legare, F., Briere, N., Stacey, D., Bourassa, H., Desroches, S., Dumont, S., ··· Roy, L. (2015). Improving Decision making On Location of Care with the frail Elderly and their caregivers (the DOLCE study): study protocol for a cluster randomized controlled trial. *Trials, 16*.

Levy, R., & Choi, J. (2013). Bayesian structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 563-623). Charlotte, NC: Information Age.

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2$×$2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods, 16*, 444-467.

Lynch, S. M. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York, NY: Springer.

McNeish, D. (2016). On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(5), 750-773.

McNeish, D., & Stapleton, L. M. (2016). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*. Advance online publication.

Muthén, B. (2010). Bayesian analysis in Mplus: A brief introduction (Incomplete draft, Version 1). Retrieved https://pdfs.semanticscholar.org/a941/cb99e52a8eafad701d4fb039134723277

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313-335.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User' s Guide* (Eighth Edition). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*, 599-620.

Newton, N. C., Teesson, M., Mather, M., Champion, K. E., Barrett, E. L., Stapinski, L., ···Slade, T. (2018). Universal cannabis outcomes from the Climate and Preventure (CAP) study: a cluster randomised controlled trial. *Substance Abuse Treatment Prevention and Policy, 13*.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling, 18*, 161–182.

Rutterford, Clare, Copas, Andrew, & Eldridge, Sandra. (2015). Methods for sample size determination in cluster randomized trials. *International Journal of Epidemiology, 44*(3), 1051-1067.

Ribeiro, D. C., Milosavljevic, S., & Abbott, J. H. (2018). Sample size estimation for cluster randomized controlled trials. *Musculoskeletal Science and Practice, 34*, 108-111.

Ryu, E. (2015). The Role of Centering for Interaction of Level 1 Variables in Multilevel Structural Equation Models. *Structural Equation Modeling-a Multidisciplinary Journal, 22*(4), 617-630.

Schoeneberger, J. A. (2015). The impact of sample size and other factors when estimating multilevel logistic models. *The Journal of Experimental Education, 84*(3).

Shen, Y., Wang, T., Gao, M., Zhu, X., Zhang, X., He, C., ···Sun, X. (2019). Effectiveness of low-cost reminder package combined with case-based health education to improve hypertensive patients' medication adherence: a clustered randomized controlled trial. *Patient Preference and Adherence, 13*, 1083-1092.

Shi, D., & Tong, X. (2017). The impact of prior information on Bayesian latent basis growth model estimation. *SAGE Open, 7*, 1–14.

Sutton, C. J., Watkins, C. L., & Dey, P. (2013). Illustrating problems faced by stroke researchers: a review of cluster-randomized controlled trials. *International Journal of Stroke, 8*(7), 566-574.

Van Breukelen, G. J. P., & Candel, M. J. (2012). Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient!. *Journal of Clinical Epidemiology, 65*(11), 1211_{2218}.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development, 85*, 842–860.

van der Putten, G.-J., Mulder, J., de Baat, C., De Visschere, L. M. J., Vanobbergen, J. N. O., & Schols, J. M. G. A. (2013). Effectiveness of supervised implementation of an oral health care guideline in care homes; a single-blinded cluster randomized controlled trial. *Clinical Oral Investigations, 17*(4), 1143-1153.

Appendix: Mplus Syntax for Bayesian Multilevel Model

```
TITLE: This is an example of Bayesian Multilevel Modeling
DATA:
FILE =data ef-nor-0.05-20-30_1.dat;
VARIABLE:
NAMES = GroupID MemID w1j x1 x2 x3 Y;
USEVAR = GroupID w1j x1 x2 x3 Y;
CLUSTER = GroupID;
WITHIN = x1 x2 x3;
BETWEEN = w1j;
ANALYSIS:
TYPE IS TWOLEVEL RANDOM;
ESTIMATOR = BAYES;
PROCESSORS = 2;
BITERATIONS = 70000 (2000);
MODEL:
%WITHIN%
b1j | Y ON x1;
b2j | Y ON x2;
b3j | Y ON x3;
%BETWEEN%
Y b1j b2j b3j;
[b1j] (b10);
[b2j] (b20);
[b3j] (b30);
Y ON w1j (b01);
b1j ON w1j (b11);
b2j ON w1j (b21);
b3j ON w1j (b31);
b1j-b3j;
MODEL PRIORS:
b01~N(1.5,0.75);
b11~N(0.12,0.06);
b21~N(-0.05,0.025);
b31~N(-0.75,0.375);
b10~N(2.3,1.15);
b20~N(-0.25,0.125);
b30~N(-0.75,0.375);
OUTPUT:
TECH1 TECH8 CINT;
```

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.